



Moderators of reputation effects in peer-to-peer online markets: a meta-analytic model selection approach

Ruohuang Jiao¹ · Wojtek Przepiorka¹ · Vincent Buskens¹

Received: 7 July 2021 / Accepted: 3 February 2022 / Published online: 7 March 2022
© The Author(s) 2022

Abstract

The effect of seller reputation on seller success in peer-to-peer online markets has been investigated in dozens of studies by means of the analysis of digital trace data. A recent meta-analysis synthesizing evidence from over a hundred studies corroborates that sellers with a better reputation sell more products at higher prices. However, the meta-analysis also shows that these reputation effects exhibit excess variation that cannot be attributed to sampling error. Moreover, there is still little consensus on how the size of a reputation effect should be interpreted and what might cause its variation. Here we use a meta-analytic model selection approach and multi-model inference on two subsets of 406 coefficient estimates to identify potential moderators of reputation effects. We identify contextual, product-related, and method-related moderators. Our results show that, among others, geographical region, product condition, sample size, and type of regression model have a bearing on the size of the reputation effect. The moderating effect of the geographical region suggests that reputation effects are substantially larger in the Chinese context than in the European or US contexts. The moderating effect of product condition—estimates based on new products are larger than estimates based on used products—is unexpected and worthwhile investigating further. The moderating effects of sample size and model type could be related to study quality. We do not find evidence for publication bias as a potential explanation for the effects of method-related moderators.

Keywords Online market · Reputation system · Reputation effect · Meta-regression · Model selection · Multi-model inference

✉ Wojtek Przepiorka
w.przepiorka@uu.nl

¹ Department of Sociology/ICS, Utrecht University, Utrecht, The Netherlands

Introduction

With the increasing popularity of online markets, more research is concerned with how reputation systems promote cooperative market exchanges. Reputation systems that are commonly employed in online market platforms collect, aggregate, and disseminate information about traders' past behaviors and the quality of their goods and services [31, 50, 60]. Trader reputation profiles are created from numeric scores (positive, neutral or negative ratings, or five-star ratings) and feedback texts (i.e. feedback messages describing the experience with certain traders and their goods and services).

Reputation systems are particularly useful for buyers, who decide which sellers to transact with but are uncertain about seller trustworthiness. In offline economic exchange, uncertainty and trust issues are often managed through networked structures through which firms establish reputations (see e.g. [44, 65], and for a review [14]). However, long-term business relations in larger business networks can hardly be established and maintained online without additional trust-building mechanisms. Reputation systems replace the role of offline networks in managing trust in online transactions. In reputation-based online markets, sellers have an incentive to be trustworthy and send back the merchandise or provide the service the buyer paid for to maintain a good reputation in favor of future business success. In addition, new sellers, who do not yet have a record of past transactions, must invest in building their reputation by reducing prices or sending other signals of their trustworthiness [24, 46, 55]. As a consequence, these sellers' reputations and their business success will be positively correlated, a phenomenon that is also known as the *reputation effect*.

Many studies have estimated the reputation effect based on digital trace data of online market transactions. Jiao, Przepiorka and Buskens [30] performed a series of meta-analyses synthesizing evidence from over a hundred such studies. Their meta-analyses corroborate the existence of reputation effects in peer-to-peer online markets. Their results provide evidence for the general relationship between seller reputation and selling performance in terms of the direction (a positive relation for positive ratings and a negative relation for negative ratings) and statistical significance. However, there is substantial heterogeneity in reputational effect sizes that cannot be explained by sampling error alone [11]. In their study, this is evidenced by high I^2 statistics, which describe the percentage of between-study variability to total variability (i.e. within and between study variability in effect sizes) (see Table 4 in [30]).

The aim of the present study is to explain the excess variation in seller reputation effects by means of the dataset created by Jiao et al. [30]. Inspired by arguments and discussions in previous literature on possible moderators of the reputational effect, we identify contextual moderators, product-related moderators and method-related moderators. For example, the market context in which the online transactions take place should be taken into consideration because traders' behavior will likely be influenced by their cultural, spatiotemporal and institutional embeddedness [8, 41]. Moreover, the types of traded products, which

range from small stamps to large motor vehicles, will also have a bearing on the size of the reputation effect. In particular, product prices, item conditions and their popularity are likely moderators of reputation effects. Finally, the methods applied across existing studies vary considerably. Even though it is possible to make reputational effect sizes comparable for the purpose of including them in meta-analyses, the variety of statistical modelling approaches will influence the estimation of the reputation effect.

We use a meta-analytic model selection approach and multi-model inference to integrate the findings from previous studies and identify potential moderators of reputation effects empirically. The model selection approach allows us to systematically consider and compare meta-regression models and determine which set of moderators contributes to the best fitting models. The multi-model inference part provides us with the relative importance of each moderator, i.e. the likelihood of each moderator to be included in a well-fitting meta-regression model. To our knowledge, we are the first to apply model selection and multi-model inference in meta-analysis to address the substantial question why reputation effects vary in size [33].

Even though we propose a few general expectations regarding potential moderators based on suggestions provided in the literature, our study is largely exploratory. Our main interest lies in determining the most influential moderators of reputational effect size within the dataset we have available. Hence, our paper contributes to the discussion on what moderators influence observable reputation effects and applies computational social science methodology to test the validity of our conjectures. Although many of the moderators that we consider in our analyses are likely correlated with variables we do not observe, this shall not prevent us from learning something from the rich dataset created by Jiao et al. [30] and discover interesting relations that could be followed up on in future research using methods that are more suitable for detecting causal relations.

Theoretical considerations

To provide explanations for the excess variation of observed reputation effects reported in Jiao et al. [30], we draw on theoretical considerations from previous literature. Potential moderators can be categorized as contextual moderators, product-related moderators and method-related moderators. In this section, we outline general expectations as to why and how these three sets of moderators might have a bearing on the size of the reputation effect. Figure 1 summarizes our considerations.

Contextual moderators

Context refers to the cultural, spatiotemporal and institutional embeddedness of online market exchanges [8, 41]. Contextual differences may lead to different attitudes towards strangers, perceptions of trustworthiness [36, 37] and propensities of leaving feedback after completed online market transactions. As a result,

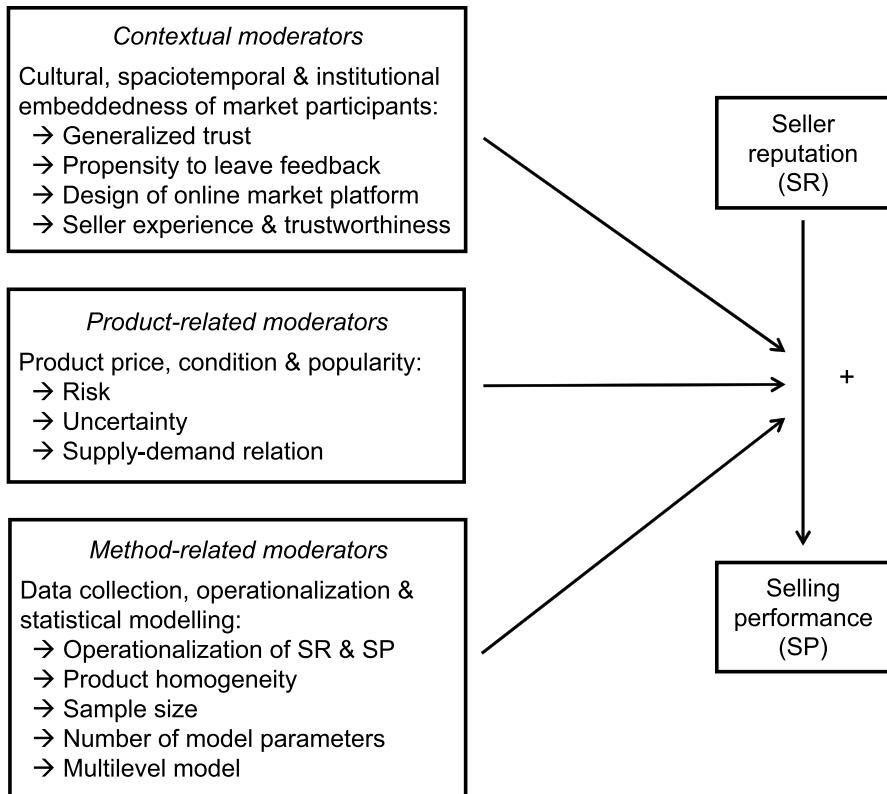


Fig. 1 Contextual, product-related, and method-related moderators of reputation effects

reputation effects may differ [53]. How do context-dependent generalized trust levels and propensities to leave feedback affect the size of the reputation effect?

Generalized trust refers to individuals' trust in strangers, i.e. people that are not part of one's family, friendship and acquaintance network [40, 49, 64]. If generalized trust is low, people may resort to other trust-building mechanisms such as reputation systems. However, because of these people's low a priori expectations of online sellers' trustworthiness, sellers with no or short records of successfully completed transactions will have to offer their items at lower prices compared to established sellers with a good reputation [30, 46]. As a consequence, the reputation effect will be larger in markets embedded in low-trust contexts.

In our analysis, we distinguish between three geographical regions: USA, China, and Europe. Theoretical arguments corroborated by empirical evidence suggest that functioning legal systems protecting property rights [9] and democratic institutions [35] promote generalized trust. This, in turn, suggests that the Chinese context will be characterized by lower levels of generalized trust and thus exhibit larger reputation effects than online market exchanges in the USA or Europe. However, this conjecture does not seem to be valid as to the level of

generalized trust, which is relatively high in China [58, 63]. The results of Steinhart's study [58] suggest that it is people's confidence in political institutions, irrespective of these institutions' suitability to protect property rights or promote democratic processes, that is positively related with generalized trust. We, therefore, refrain from stating expectations regarding the moderating effect of the geographical region (via generalized trust) on reputation effects.

Not leaving truthful feedback after a bad experience with an online transaction or leaving overly positive feedback after a mediocre experience can lead to a positive bias in seller reputations. As a consequence, untrustworthy sellers selling low-quality items will be identified with a delay or not at all, which will make it more attractive for these sellers to enter the market. Because, as a result, the likelihood of encountering an untrustworthy seller will be higher, buyers will demand larger discounts when dealing with sellers without an established reputation record. This, in turn, will lead to a larger reputation effect [30, 46].

In as far as people differ in their propensity to leave truthful feedback after completed online market transactions, contextual factors can also affect the magnitude of the reputation effect. For example, in some cultures there are norms proscribing overly critical feedback whereas in other cultures there are norms proscribing overly praising feedback. Zhang and Huang [69] state that in China buyers rarely give negative ratings, and in case of a negative experience with a seller, buyers tend to resort to neutral ratings. However, there is also evidence that people are reluctant to give negative feedback in the US context and rather refrain from giving feedback after a bad experience [18, 42]. Relatedly, results from the Global Preference Survey [25] indicate that people's intentions to positively reciprocate good deeds and to negatively reciprocate misdeeds may not be so different in China and the USA, whereas in Europe people on average exhibit more variation in these intentions across countries. We, therefore, refrain from stating expectations regarding the moderating effect of the geographical region (via feedback behavior) on reputation effects.

Differences in design features of online market platforms and reputation systems may instigate the emergence of different rating conventions through which buyers might perceive seller reputations differently [1]. For example, the possibility to leave feedback may be used to establish seller trustworthiness, but in a two-sided rating system, where buyers can rate sellers and vice versa, ratings can be used as a means to positively or negatively reciprocate one's trading partner's positive or negative rating, respectively [10, 19, 21]. As a consequence of the threat to retaliate a negative rating with a negative rating that can be upheld in two-sided rating systems, seller reputations will be positively biased. Hence, buyers may trust seller reputation information in such systems less, so that the reputation effect will be smaller [62]. However, since information on how truthful ratings in different online market platforms are is not usually available and reported, we account for the market platform as a potential moderator but do not state any expectations as to the direction of its effect on the reputation effect.

Since its first implementation on eBay, online reputation systems have gone through changes with the aim to make reputation systems more effective in keeping untrustworthy sellers out of business (e.g., [10, 21, 51]). Therefore, over time, buyers may have become more aware of the effectiveness of reputation systems. If

over time, the average seller quality has indeed become higher as a consequence of the improvements in reputation systems, we can expect the reputation effect to become smaller; if more sellers are trustworthy, buyers need to rely less on information about the seller reputations when making their buying decisions. In our analysis, we will account for the time of data collection as a moderator of the reputation effect and expect its effect to be negative. However, the argument that, over time, the reputation effect will become smaller because average seller quality increases, hinges on the assumption that the number of new sellers entering the market is relatively low. A substantial number of sellers that enter the market anew may positively affect the size of the reputation effect because these sellers still need to build their reputation by offering price discounts [46, 55]. However, information on the average experience of sellers in a particular online market platform is not usually reported and therefore not available to be included in our analyses. Figure 1 summarizes the contextual moderators of reputation effects in peer-to-peer online markets.

Product-related moderators

The size of the reputation effect may systematically vary with product features because of the information asymmetries buyers face in online transactions and the risks and uncertainties that result from such information asymmetries [3, 20]. The risk that a buyer takes in an online market transaction depends on the probability of the seller being trustworthy and the price of the traded product. Since more expensive products exhibit a higher risk, a buyer might be willing to pay a higher price to a seller with a better reputation to reduce the probability of being cheated [67]. Hence, the higher the price of the traded product, the larger will be the reputation effect.

The uncertainty buyers face in online transactions also stems from the uncertainty regarding product quality. In peer-to-peer online markets both new and used products are sold. While uncertainty about product quality is low for new products, it will be high for used products. For used products, buyers' expectations are formed based on how sellers describe and present these products and, therefore, eventually, on how trustworthy buyers expect these sellers to be. We thus expect that for used products or for products of unknown condition, seller reputation information will play a more important role for buyers. As a result, the reputation effect will be larger for used products or products of unknown condition than for new products [19].

Finally, the extent of market clearing will influence the size of the reputation effect. If the market for a certain product experiences higher demand than supply (e.g., because of the product's newness and popularity), buyers may be willing to take higher risks and pay less attention to seller reputations and other signs of seller trustworthiness [22, 45]. However, if supply is higher than demand, sellers will experience more competition among each other and sellers with a lower reputation may need to accept lower prices for their products to be chosen by buyers [26]. Therefore, we expect that in markets in which demand exceeds supply, the seller reputation effect will be smaller than in markets in which supply exceeds demand.

Method-related moderators

Various types of operationalizations of seller reputation have been used among the studies included in our meta-analyses. For example, the reputation score (the number of positive ratings a seller received minus the number of negative ratings), the number of positive ratings, the percentage of positive ratings (the percentage of positive ratings among all received ratings), etc., have been used as measures of seller reputation. In most online market platforms, the number and percentage of positive ratings, the reputation score and the number of negative ratings are presented on seller profile pages, but it is researchers' decisions which of these measures are used in their analyses. Although the different operationalizations of seller reputation are highly correlated with each other [71], the size of the reputation effect may still depend on which one is used in statistical data analysis. For example, the reputation effect may be lower if it is estimated based on the percentage of positive ratings because the variance of this variable appears to be small [4, 67]. In many online market platforms, most sellers have very high percentages of positive ratings (98% and higher). This may result from the default rating set by the platform [48], or from sellers trying to prevent any non-positive ratings (even with inappropriate means) [17]. Thus, the small range of the percent-of-positive-ratings variable may provide too little leverage to identify a sizable reputation effect. Another reason for why using the percentage of positive ratings might result in a lower reputation effect is that it is not a valid measure of seller trustworthiness. Sellers with 50 transactions and sellers with 1000 transactions alike can have 99% positive ratings but the latter can be considered as more trustworthy than the former. This is because a good reputation must be costly to acquire and therefore 1000 mostly positive ratings are a stronger sign of trustworthiness than 50 mostly positive ratings [47]. Therefore, we assume that the size of the reputation effect will be smaller if sellers' positive (or negative) ratings are measured in terms of percentages rather than absolute numbers.

Moreover, whether the variables measuring seller reputations and product prices are log-transformed or untransformed may also influence the size of the reputation effect. If the relation between seller reputation and selling performance indeed is non-linear (e.g., increasing at a decreasing rate; [46]), a linear model might produce a downward biased reputation effect. We thus expect models with log-transformed seller reputation and selling performance to produce larger reputation effects. To better understand how reputation effects depend on such method-related moderators, the types of operationalizations of seller reputation and selling performance are included in the exploratory model selection process.

Whether the traded products are homogeneous within each included dataset is a second aspect of methodological concern that is likely to affect the size of the reputation effect. Although this might appear to be a product-related moderator (see above), we include product homogeneity as a method-related moderator; it is researchers' decisions how homogeneous the products are they collect data on and analyze in their studies. Even for the same category of products (e.g., mobile phones), some researchers choose a specific category to collect data on (e.g., Eddhir [23] collected data on unlocked iPhone 3G), whereas other researchers choose a more general product category (e.g., Zhou [70] collected data on mobile phones)

and provide no further information on what the product category comprises. We assume that there is a larger variation in observed reputation effects if a dataset of heterogeneous products is used, and we account for product heterogeneity in our meta-analyses. However, we do not have any expectations as to the direction of its effect on the reputation effect.

Finally, we expect that the way in which the statistical models are constructed with which reputation effects are estimated will have a bearing on the size of the reputation effect. We identify three aspects of statistical model building: (1) whether it is a multi-level model, (2) the number of observations or clusters and (3) the number of parameters. As for the type of model, a sizable proportion of models account for the repeated observations on same sellers by fitting multi-level regression models. In particular, these models estimate cluster-robust standard errors of coefficient estimates to account for the dependence of observations stemming from the same seller. The number of observations, clusters and parameters influence the calculation of reputational effect sizes through the *degrees of freedom* ($df = N - k - 1$, where N indicates the number of observations or number of clusters and k indicates the number of parameters). Models with more cases are more likely to detect smaller reputation effects. However, this does not imply that the moderating effect of the number of cases will be negative. Models with cluster-robust standard errors tend to be conservative in terms of estimating significance, but do not affect the effect size. Thus, we do not state any expectations as to the net direction of the effect of a number of observations, clusters and parameters on the size of the reputation effect.

Control variables

Apart from the moderators that we identify above based on theoretical considerations and evidence from empirical research, we include a few control variables to capture attributes that are commonly reported and appear to be relevant to explain the between-study or between-dataset differences in reputation effects. The control variables are the publication status of the study (published in English vs published in other language vs not published) and the type of transaction captured in a dataset (auction vs fixed price offer).

Data and methods

The meta-analytic dataset was collected until October 2021. It includes 406 estimates of reputation effects (i.e. effects of seller reputation on selling performance), estimated with 202 different datasets, and reported in 125 empirical papers (also see [30]). The literature search started with two previous meta-analytic studies [34, 54] and was extended with literature searches to include more recent studies and studies written in other languages than English.¹ This search process resulted in 205

¹ Compared to the two previous meta-analyses, the current study not only updates and extends the data (Liu and colleagues [34] include 42 articles and Schlägel [54] includes 58 articles), but also extends our insights by using different methodology. The previous studies used combined significance tests and sign tests as the meta-analytic approaches, which only consider the sign of reputation effects. In our study, we

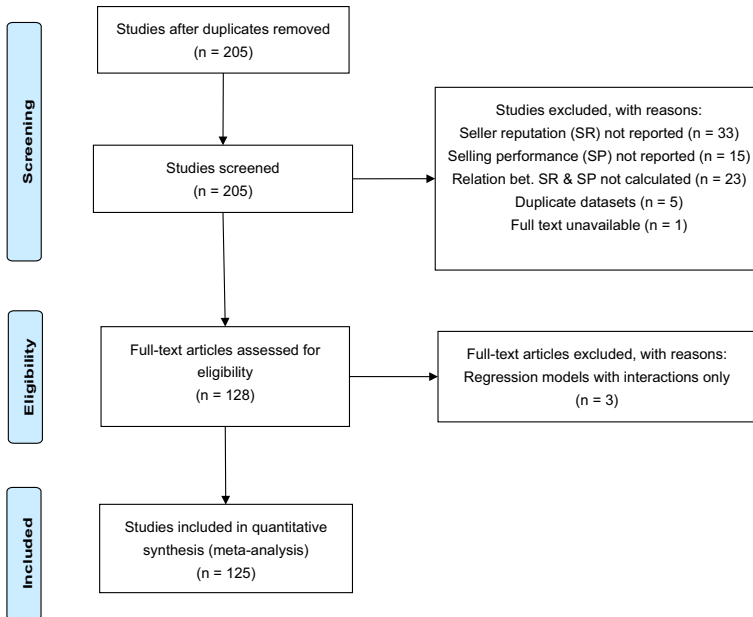


Fig. 2 PRISMA flow diagram of literature search and inclusion criteria [39]. The complete list of the 205 articles and the reasons for exclusion (if excluded) is provided in Online Resource 1, and the complete reference list is provided in Online Resource 2

relevant research articles written in English, Chinese or German. The PRISMA flow diagram presented in Figure 2 outlines the criteria we used for the inclusion of studies for our meta-analyses. Further details on search criteria, study inclusion criteria and the calculation of effect sizes are described in Jiao et al. [30].

Table 1 provides basic descriptive statistics of reputational effect sizes² across potential moderators at the study level and dataset level. Descriptive statistics are provided separately for the three different types of seller reputation information: reputation score, positive ratings and negative ratings. What is apparent from Table 1 is the large extent of variance in reputational effect sizes within categories of potential moderator variables. In what follows, we will describe each moderator variable in detail.

Footnote 1 (continued)

transform the estimates of the reputation effects into comparable effect sizes. This allows us to assess the variation of these effect sizes and identify potential moderators of the reputation effect.

² The effect sizes were calculated as Pearson correlations from bivariate relationships with $r = \rho$, or partial correlations from multiple regression models with $r = \frac{t}{\sqrt{t^2 + df}}$, depending on the information reported in included studies. For more details, see [30].

Moderator variables

Region indicates the country or region where the dataset was collected. It reflects the cultural difference among datasets. As is shown in Table 1, most datasets were collected in the USA and China. Datasets in the category “Europe” were collected in France, Poland, Germany, Finland and Switzerland.

Market platform indicates on which market platform the dataset was collected, taken to explain the institutional differences among datasets. As is shown in Table 1, most datasets were collected from eBay (the largest peer-to-peer online market in the world) and Taobao (the largest peer-to-peer online market in China). Category “Other” includes the following platforms: Yahoo! (in USA and China), Eachnet (China), Priceminister (France), Allegro (Poland), Huuto (Finland), Bizerate (USA), Ricardo (Switzerland), Bonanza (USA) and Silkroad (a platform focusing on illegal drugs).³

Year of data collection is a continuous variable, indicating the year in which the collection of the dataset started. The variable ranges from 1998 to 2019. It is included in our analysis to capture the influence of the time period on the size of reputation effects.

Condition indicates whether the products comprised in a dataset were marketed new or used. The category “unknown” indicates that this information is not mentioned in the paper or the dataset includes both new and used products without differentiating between them.

Average price is a continuous variable reporting the mean price of products in the collected dataset. Given datasets were collected in various currencies, the mean prices are converted in US Dollars with Purchase Power Parities (PPP)⁴ for comparability across countries with different price levels. The mean prices were mostly reported in the data description section of included papers, but for 55 of the 202 datasets the mean price was not reported. To reduce the number of missing values, we estimate the mean price in datasets with missing values based on product mean prices from similar datasets.⁵ In this way, we obtained for 17 of the 55 datasets an estimated product price. In our analysis, we use the log-transformed average price.

Rate of market clearing indicates how large the proportion of sold products contained in a dataset was. This variable is a proxy for the popularity of the product. However, for 144 out of 202 datasets this information is not reported. To avoid too many missing values, this variable is not included in our analysis.

Log-transformed SR is a dichotomous variable that indicates whether the variable of seller reputation was log-transformed or not.

³ Jian et al. [29] collect data from an unspecified platform, and the dataset from Snijders and Zijdemans [57] comes from four different platforms.

⁴ <https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm>.

⁵ The estimation is based on similar datasets with reported product mean prices. For instance, datasets with similar types of products (e.g., SD card and U disk), year of data collection (within 5 years), and same usage condition (i.e. new, used or unknown). If no similar dataset is found, it is estimated with an average value of price within the same product category (e.g., mobile phones).

Table 1 Summary of effect sizes (*ES*) for each category of seller reputation

| ES | | Reputation score | | | Positive ratings | | | Negative ratings | | |
|--------------------------------------------------|----|------------------|------|----|------------------|------|----|------------------|------|----|
| | | <i>n</i> | Mean | SD | <i>n</i> | Mean | SD | <i>n</i> | Mean | SD |
| <i>Contextual moderators</i> | | | | | | | | | | |
| Region | | | | | | | | | | |
| United States | 88 | 0.03 | 0.18 | 73 | 0.19 | 0.23 | 57 | -0.07 | 0.17 | |
| China | 56 | 0.08 | 0.23 | 46 | 0.15 | 0.19 | 29 | -0.10 | 0.24 | |
| Europe | 13 | 0.02 | 0.07 | 23 | 0.12 | 0.14 | 20 | -0.10 | 0.10 | |
| Platform | | | | | | | | | | |
| eBay | 85 | 0.02 | 0.17 | 80 | 0.18 | 0.22 | 69 | -0.08 | 0.16 | |
| Taobao | 49 | 0.06 | 0.24 | 44 | 0.14 | 0.20 | 22 | -0.11 | 0.26 | |
| Other | 23 | 0.09 | 0.16 | 18 | 0.16 | 0.15 | 15 | -0.07 | 0.13 | |
| Year of data collection | | | | | | | | | | |
| 1998–2002 | 61 | 0.03 | 0.15 | 32 | 0.28 | 0.26 | 36 | -0.04 | 0.18 | |
| 2003–2007 | 27 | 0.03 | 0.18 | 55 | 0.12 | 0.18 | 41 | -0.09 | 0.13 | |
| 2008–2012 | 42 | 0.02 | 0.20 | 43 | 0.15 | 0.17 | 27 | -0.11 | 0.22 | |
| 2013–2019 | 27 | 0.14 | 0.25 | 12 | 0.08 | 0.07 | 2 | -0.28 | 0.35 | |
| <i>Product-related moderators</i> | | | | | | | | | | |
| Condition | | | | | | | | | | |
| New | 60 | 0.07 | 0.24 | 83 | 0.20 | 0.22 | 69 | -0.07 | 0.20 | |
| Used | 25 | -0.02 | 0.17 | 10 | 0.01 | 0.12 | 8 | -0.09 | 0.14 | |
| Unknown | 72 | 0.05 | 0.15 | 49 | 0.12 | 0.17 | 29 | -0.10 | 0.12 | |
| Average price (in US dollar, in 25% percentiles) | | | | | | | | | | |
| 2.86–26.63 | 41 | 0.06 | 0.16 | 25 | 0.15 | 0.14 | 19 | -0.02 | 0.13 | |
| 26.63–156.12 | 43 | 0.01 | 0.21 | 24 | 0.18 | 0.19 | 16 | -0.07 | 0.16 | |
| 156.12–353.6 | 16 | 0.03 | 0.12 | 36 | 0.16 | 0.20 | 33 | -0.07 | 0.15 | |

Table 1 (continued)

| ES | | Reputation score | | | Positive ratings | | | Negative ratings | | |
|----------------------------------------------|----------|------------------|------|----------|------------------|------|----------|------------------|------|--|
| | <i>n</i> | Mean | SD | <i>n</i> | Mean | SD | <i>n</i> | Mean | SD | |
| 353.6–19,342.4 | 24 | 0.08 | 0.23 | 34 | 0.19 | 0.28 | 24 | -0.10 | 0.21 | |
| Missing | 33 | 0.06 | 0.21 | 23 | 0.12 | 0.15 | 14 | -0.15 | 0.24 | |
| Rate of market clearing (in 25% percentiles) | | | | | | | | | | |
| 0–49.34% | 11 | 0.05 | 0.09 | 13 | 0.12 | 0.15 | 12 | -0.08 | 0.09 | |
| 49.34–64.3% | 15 | 0.04 | 0.09 | 11 | 0.08 | 0.08 | 9 | -0.10 | 0.08 | |
| 64.3–82.58% | 20 | 0.04 | 0.09 | 8 | 0.09 | 0.14 | 6 | -0.10 | 0.18 | |
| 82.58–98% | 8 | -0.02 | 0.10 | 16 | 0.09 | 0.09 | 11 | -0.10 | 0.08 | |
| Missing | 103 | 0.05 | 0.23 | 94 | 0.20 | 0.23 | 68 | -0.07 | 0.21 | |
| <i>Method-related moderators</i> | | | | | | | | | | |
| Log-transformed/percentage SR | | | | | | | | | | |
| Log-trans. | 75 | 0.05 | 0.22 | 78 | 0.22 | 0.22 | 73 | -0.07 | 0.19 | |
| Percentage | n/a | n/a | n/a | 50 | 0.11 | 0.15 | 13 | -0.04 | 0.08 | |
| Untrans. | 82 | 0.04 | 0.16 | 14 | 0.03 | 0.17 | 20 | -0.15 | 0.16 | |
| Log-transformed SP | | | | | | | | | | |
| Log-trans. | 52 | 0.04 | 0.25 | 52 | 0.17 | 0.17 | 32 | -0.14 | 0.19 | |
| Untrans. | 105 | 0.05 | 0.16 | 90 | 0.16 | 0.22 | 74 | -0.06 | 0.17 | |
| Product homogeneity | | | | | | | | | | |
| Homogeneous | 136 | 0.04 | 0.20 | 126 | 0.17 | 0.21 | 99 | -0.08 | 0.18 | |
| Heterogeneous | 21 | 0.06 | 0.11 | 16 | 0.09 | 0.09 | 7 | -0.10 | 0.14 | |

Table 1 (continued)

| ES | | Reputation score | | | Positive ratings | | | Negative ratings | | |
|------------------------------------------------------|----------|------------------|------|----------|------------------|------|----------|------------------|------|--|
| | <i>n</i> | Mean | SD | <i>n</i> | Mean | SD | <i>n</i> | Mean | SD | |
| Multilevel | | | | | | | | | | |
| Multilevel | 18 | 0.05 | 0.17 | 25 | 0.12 | 0.14 | 19 | -0.12 | 0.13 | |
| Other | 139 | 0.04 | 0.20 | 117 | 0.17 | 0.22 | 87 | -0.08 | 0.19 | |
| Number of observations <i>N</i> (in 25% percentiles) | | | | | | | | | | |
| 14–130 | 30 | -0.06 | 0.30 | 40 | 0.34 | 0.25 | 33 | -0.08 | 0.25 | |
| 130–389 | 40 | 0.12 | 0.21 | 33 | 0.12 | 0.14 | 27 | -0.04 | 0.14 | |
| 389–2433 | 43 | 0.04 | 0.11 | 35 | 0.09 | 0.14 | 24 | -0.11 | 0.14 | |
| 2433–3,051,759 | 44 | 0.05 | 0.11 | 34 | 0.07 | 0.09 | 22 | -0.09 | 0.12 | |
| Number of parameters <i>k</i> (in 25% percentiles) | | | | | | | | | | |
| 1–3 | 21 | 0.09 | 0.30 | 52 | 0.23 | 0.27 | 41 | -0.02 | 0.20 | |
| 3–6 | 47 | -0.01 | 0.23 | 27 | 0.15 | 0.17 | 14 | -0.23 | 0.22 | |
| 6–14 | 47 | 0.06 | 0.10 | 29 | 0.13 | 0.15 | 30 | -0.10 | 0.13 | |
| 14–348 | 42 | 0.07 | 0.15 | 34 | 0.09 | 0.09 | 21 | -0.08 | 0.06 | |
| <i>Controls</i> | | | | | | | | | | |
| Publication status | | | | | | | | | | |
| English journal | 108 | 0.04 | 0.17 | 83 | 0.17 | 0.22 | 81 | -0.06 | 0.16 | |
| Local journal | 24 | 0.14 | 0.19 | 30 | 0.11 | 0.11 | 10 | -0.09 | 0.20 | |
| Other | 25 | -0.02 | 0.25 | 29 | 0.19 | 0.23 | 15 | -0.16 | 0.25 | |
| Transaction type | | | | | | | | | | |
| Auction | 99 | 0.04 | 0.14 | 82 | 0.17 | 0.21 | 77 | -0.07 | 0.16 | |

Table 1 (continued)

| ES | | Reputation score | | Positive ratings | | Negative ratings | |
|-------------|----------|------------------|------|------------------|------|------------------|-------|
| | <i>n</i> | Mean | SD | <i>n</i> | Mean | <i>n</i> | Mean |
| Fixed price | 54 | 9.97 | 0.24 | 51 | 0.14 | 25 | -0.13 |
| Unknown | 4 | -0.22 | 0.39 | 9 | 0.25 | 4 | -0.12 |
| | | | | | | | SD |
| | | | | | | | 0.24 |
| | | | | | | | 0.06 |

For each category of seller reputation, the number of effect sizes that are included in the meta-analytic dataset, their mean and standard deviation are reported. Year of data collection, average price, rate of market clearing, number of observations and number of parameters are treated as continuous variables in the data analysis. SR stands for seller reputation and SP for selling performance

Percentage SR is a dichotomous variable that indicates whether seller reputation is measured in percentages (e.g., percent of positive ratings) or in absolute terms. This variable only applies to positive and negative ratings as measures of seller reputation but not to reputation scores, which is the number of positive ratings minus the number of negative ratings.

Log-transformed SP is a dichotomous variable that indicates whether the variable of selling performance was log-transformed or not.

Product homogeneity is a dichotomous variable indicating whether the products contained in a dataset are homogeneous or heterogeneous.

Multilevel is a dichotomous variable that indicates whether data analysis includes and accounts for clustered observations by using multilevel techniques.

N is the number of observations in each model that is included. In our analysis, we use the log-transformed number of observations.

Parameters is the number of parameters that was included in each model. In our analysis, we use the log-transformed number of parameters.

Publication status is collected at the study level and indicates whether the paper reporting the results has been published. The publication types include International/English journals, local journals and other, i.e. book chapters, conference/workshop presentations, dissertation/thesis, working paper, and unpublished. To simplify the variable, the publication status is treated as a nominal variable during data analysis. That is, papers are treated as “published in English journals”, “published in local journals” and “other (unpublished)”.

Transaction type reports whether the collected dataset contains auctions or fixed-price transactions. It is categorized as “unknown” if this information is not reported in the paper or the dataset contains mixed types of transactions.

Reproduction of previous findings with a meta-regression model

Jiao et al. [30] ran 12 separate meta-analyses, one for each combination of seller reputation (reputation score, number of positive ratings, number of negative ratings) and selling performance (final price, price ratio, selling probability, selling volume), using 378 effect sizes reported in 107 empirical studies. Their dataset has been updated in the meantime and now comprises 406 effect sizes reported in 125 studies. Here we reproduce their results by fitting a saturated, random-effects meta-regression model with the updated sample of 406 reputational effect sizes. In this model, the type of seller reputation and the type of selling performance are the only explanatory variables and are fully interacted with each other. The results are presented in Table 2 and correspond to the findings reported by Jiao et al. [30].

In general, seller reputation has a small but significant effect on selling performance. The overall effects of reputation score and positive ratings are positive, and the overall effects of negative ratings are negative. Moreover, in absolute terms, the overall effects of positive ratings are larger than both the overall effects of reputation score and negative ratings. Three out of twelve overall effect sizes are statistically insignificant: reputation score ($p=0.19$) and negative ratings ($p=0.10$) on price ratio and negative ratings on selling volume ($p=0.36$).

The significant Q-statistic for moderators ($QM_{(df = 12)} = 160.75, p < 0.001$) suggest that the inclusion of all interactions of type of seller reputation with the type of selling performance, which corresponds to the twelve sub-group meta-analyses conducted by Jiao et al. [30], explains a substantial proportion of the variation in reputational effect sizes. However, the significant Q-statistic for residual heterogeneity ($QE_{(df = 394)} = 17,917.34, p < 0.001$) suggests that further exploration with potential moderators is likely to be worthwhile because the amount of residual heterogeneity is very high [11, 27]. Next, we explain in more detail our data analysis strategy.

Although the total number of 406 reputational effect sizes used in the meta-regression reported in Table 2 appears to be large, the number of cases will be quickly diminished in sub-group analyses or meta-regressions that include additional moderators. In order to test the influence of the potential moderators listed earlier, we will focus on two homogeneous subsets of the dataset. To create these subsets, we first pool cases that use final price or price ratio as outcome variables because they are both measures of product prices. While the final price is the absolute selling price, the price ratio is the relative selling price compared to a reference value such as a book value provided by a third party or the average price of similar products in the dataset. Moreover, we pool cases that use selling probability or selling volume as outcome variable because they both capture the number of sales; a higher selling probability would imply a higher selling volume within a certain time period. We also pool cases that use reputation score (number of positive ratings minus number of negative ratings) or number of positive ratings to operationalize seller reputations because the two variables are highly correlated and both exhibit positive effects on selling performance (Table 2). Next, we create two subsets of the original dataset for further analyses. Subset 1 comprises the 183 cases that use the final price or price ratio to operationalize selling performance and reputation score or number of positive ratings to operationalize seller reputation. Subset 2 comprises the 117 cases that use selling probability or selling volume to operationalize selling performance and reputation score or number of positive ratings to operationalize seller reputation. In our analyses, we include dummy variables to control for the type of selling performance and seller reputation operationalizations used within each subset (e.g., DV is final price vs price ratio). Unfortunately, due to the small number of cases, we are unable to use the subset of cases that used negative ratings to operationalize seller reputation in our analyses.

Model selection and multi-model inference

To investigate how potential moderators affect reputation effects, we take an information-theoretic approach and apply model selection analysis and multi-model inference to our rich meta-analytic dataset. Model selection analysis examines several competing models simultaneously to identify the best set of models via information criteria such as the Akaike information criterion (AIC) [13] and model weights (aka Akaike weights) that indicate the probability that a model is a best-fitting model. In this way, it is possible to uncover (in statistical terms) the model that explains the dataset best given different combinations of moderator variables. Furthermore, we

Table 2 Reproduction of Table 4 in Jiao et al. [30] with an updated dataset ($n=406$) and a saturated, random-effects meta-regression model

| Relation | Coef | <i>p</i> -value | 95% CI | <i>n</i> | <i>Z</i> * |
|----------------------------------------|----------|-----------------|----------------|----------|------------|
| Final price × reputation score | 0.04* | 0.02 | [0.01, 0.08] | 75 | -0.65 |
| Final price × positive ratings | 0.11*** | <0.001 | [0.07, 0.15] | 56 | 2.58** |
| Final price × negative ratings | -0.09*** | <0.001 | [-0.14, -0.05] | 47 | -2.55* |
| Price ratio × reputation score | 0.07 | 0.19 | [-0.04, 0.18] | 17 | 0.43 |
| Price ratio × positive ratings | 0.27*** | <0.001 | [0.18, 0.35] | 35 | 5.10*** |
| Price ratio × negative ratings | -0.09 | 0.10 | [-0.19, 0.02] | 26 | 2.51** |
| Selling probability × reputation score | 0.04** | 0.003 | [0.01, 0.07] | 26 | -2.25** |
| Selling probability × positive ratings | 0.07*** | <0.001 | [0.04, 0.10] | 19 | 0.05 |
| Selling probability × negative ratings | -0.05** | 0.004 | [-0.09, -0.02] | 16 | -1.73 |
| Selling volume × reputation score | 0.10** | 0.002 | [0.04, 0.16] | 39 | -2.19** |
| Selling volume × positive ratings | 0.14*** | <0.001 | [0.08, 0.21] | 33 | 4.66*** |
| Selling volume × negative ratings | -0.05 | 0.36 | [-0.14, 0.05] | 17 | -3.01** |

Test of moderators (coefficients 1:12): $QM(df=12) = 160.75, p < 0.001$. Test for residual heterogeneity: $QE(df=394) = 17,917.34, p < 0.001$

*Egger's regression test for funnel plot asymmetry, as the assessment of publication bias. All the funnel plots are reported in Online Resource 3

perform multi-model inference to better evaluate the importance of each included moderator. We report the relative importance value of each moderator, which is the sum of Akaike weights of all models that include the moderator. Hence, a moderator that is included in more models with larger weights will receive a higher importance value. The advantage of multi-model inference is that it reduces the risk of selecting one of the less probable models by chance, because the relative importance of all moderators is listed [16]. With this approach, we can learn which moderators play an important role in explaining the variance in reputational effect sizes.

The combination of model selection and multi-model inference with meta-analyses has been used in linguistics [38], biology [52], biogeoscience [32], psychiatry [28] and ecology [15]. This approach is still rarely used in the social sciences. Young and Holsteen [68] introduced multi-model analysis as a methodological application in sociology to examine the choice of controls and check the robustness of empirical results with regard to model specification. Here we apply the combination of these methods to address a sociologically relevant question: What affects the size of the seller reputation effect in peer-to-peer online markets? Our analyses are conducted in R using the “MuMIn” [7] and “metafor” [66] packages.

We start our analyses by applying model selection to Subset 1 and Subset 2. We examine the fit and plausibility of random effects, meta-regression models with all possible combinations of moderators and select three best-fitting models with Subset 1 (Table 4) and one best-fitting model with Subset 2 (Table 5) for interpretation. To better grasp the importance of the various moderators, we apply multi-model inference and report the relative importance of each moderator in the last column of the two regression tables. Because Egger’s test (see Table 4 in [30]) suggests that there might be publication bias in our set of studies, we do a robustness check for publication bias by adding SE to the best-fitting meta-regression models [6, 43]. We do not find any significant effects of SE (or SE²) suggesting limited evidence for publication bias. Detailed results are reported in Online Resource 4.

Results

Table 3 presents descriptive information based on the entire sample of all moderators to be considered in our analyses. *Average price* is the only variable with missing values.

The results of the model selection analysis for Subset 1 are presented in Table 4. The best-fitting models are ranked by Akaike information criterion corrected for small sample size (AICc) [2, 9]. AICc is calculated by

$$AICc = -2\log(L) + 2k + \frac{2k(k+1)}{n-k-1} \quad (1)$$

with the number of model parameters (k), the maximum likelihood estimate for the model (L) and the sample size (n). The model with the lowest AICc is considered

Table 3 Descriptive information of moderators ($n=406$)

| Variable | Type | Mean | SD |
|-----------------------------------|-------------|-----------|-----------|
| <i>Contextual moderators</i> | | | |
| Region | Nominal | | |
| United States | | 0.54 | 0.50 |
| China | | 0.32 | 0.46 |
| Other | | 0.14 | 0.35 |
| Platform | Nominal | | |
| eBay | | 0.58 | 0.49 |
| Taobao | | 0.28 | 0.44 |
| Other | | 0.14 | 0.29 |
| Year of data | Continuous | 2006 | 5.03 |
| <i>Product-related moderators</i> | | | |
| Condition | Nominal | | |
| New | | 0.52 | 0.50 |
| Used | | 0.11 | 0.31 |
| Unknown | | 0.37 | 0.48 |
| Avg. price ^{a*} | Continuous | 589.30 | 2094.36 |
| <i>Method-related moderators</i> | | | |
| Log-transformed SR | Dichotomous | 0.56 | 0.50 |
| Percentage SR | Dichotomous | 0.16 | 0.37 |
| Log-transformed SP | Dichotomous | 0.33 | 0.47 |
| Product homo | Dichotomous | 0.89 | 0.31 |
| Multilevel | Dichotomous | 0.15 | 0.36 |
| N^* | Continuous | 47,438.74 | 245,755.9 |
| Parameter* | Continuous | 23.33 | 109.87 |
| <i>Controls</i> | | | |
| Published | Nominal | | |
| English journal | | 0.67 | 0.47 |
| Local journal | | 0.16 | 0.36 |
| Other | | 0.17 | 0.38 |
| Transaction type | Nominal | | |
| Auction | | 0.64 | 0.48 |
| Fixed price | | 0.32 | 0.46 |
| Unknown | | 0.04 | 0.20 |

^a70 missing values on the average item price

*Variables are log-transformed in the model selection analysis

the best fitting model. Table 4 presents the best fitting model (M1) as well as the two models (M2 and M3) that are less than 1.5 AICc units away from M1. The weight information listed at the bottom of each column is the Akaike weight of each model.

Table 4 Best-fitting models and relative importance of moderators for Subset 1 focusing on effects of measures of positive reputation on selling price ($n = 183$)

| | M1 | | M2 | | M3 | | Relative importance |
|-----------------------------------|----------|------|----------|------|----------|------|---------------------|
| | Coef | SE | Coef | SE | Coef | SE | |
| Const | 0.39*** | 0.05 | 0.39*** | 0.05 | 0.39*** | 0.05 | |
| <i>Contextual moderators</i> | | | | | | | |
| Region | excl | | Excl | | excl | | 0.13 |
| Platform | excl | | | | excl | | 0.39 |
| eBay | | | (ref.) | | | | |
| Taobao | | | -0.04 | 0.04 | | | |
| Other | | | 0.05 | 0.04 | | | |
| Year of data | excl | | excl | | excl | | 0.26 |
| <i>Product-related moderators</i> | | | | | | | |
| Condition | | | | | | | |
| New | (ref.) | | (ref.) | | (ref.) | | 0.99 |
| Used | -0.17*** | 0.04 | -0.17*** | 0.04 | -0.19*** | 0.04 | |
| Unknown | -0.02 | 0.03 | -0.02 | 0.03 | -0.03 | 0.03 | |
| Avg. price | excl | | excl | | excl | | <0.01 |
| <i>Method-related moderators</i> | | | | | | | |
| Log-transformed SR | excl | | excl | | excl | | 0.32 |
| Percentage SR | excl | | excl | | 0.06 | 0.04 | 0.41 |
| Log-transformed SP | excl | | excl | | excl | | 0.25 |
| Product homo | excl | | excl | | excl | | 0.27 |
| Multilevel | 0.14** | 0.04 | 0.12** | 0.04 | 0.15*** | 0.04 | 0.95 |
| Log (N) | -0.03*** | 0.01 | -0.03*** | 0.01 | -0.03*** | 0.01 | 0.98 |
| Log (parameter) | excl | | excl | | excl | | 0.39 |
| <i>Controls</i> | | | | | | | |
| Published | excl | | excl | | excl | | 0.25 |
| Transaction type | excl | | excl | | excl | | 0.12 |
| DV final price | -0.10** | 0.03 | -0.08** | 0.03 | -0.10*** | 0.03 | 0.97 |
| IV reputation score | -0.05 | 0.03 | -0.06* | 0.03 | excl | | 0.56 |
| n | 183 | | 183 | | 183 | | |
| Log likelihood | 39.72 | | 41.35 | | 39.09 | | |
| AICc | -62.6 | | -61.4 | | -61.3 | | |
| Weight | 0.033 | | 0.018 | | 0.018 | | |

The sum of Akaike weights of all possible models is 1 [13, 61]. The Akaike weight of M1 indicates that there is a 3.3% chance that M1 is the best model for explaining the data, and that is the highest chance among all fitted models.

Model M1 includes usage conditions, a product-related moderator. The detected reputation effect is significantly smaller for used products than new products (coef. = -0.17, $p < 0.001$), which is contrary to our expectation. Because of the

higher uncertainty regarding the condition of used products, we expected the reputation effect to be larger for used products than for new products.

M1 also includes method-related moderators of the reputation effect. If the model accounts for clustered data, the reputation effects are significantly larger (coef. = 0.14, $p = 0.002$). The number of observations (coef. = -0.03, $p < 0.001$) shows a significantly negative effect, which indicates that models with larger samples exhibit smaller reputation effects.

As for control variables, we find a significant distinction between the two types of dependent and independent variables used in models included in our meta-regressions. The effect of seller reputation on final price is smaller than on price ratios (coef. = -0.10, $p = 0.002$). Besides, reputation effects appear to be smaller when the independent variable is reputation score instead of positive ratings (coef. = -0.05, $p = 0.05$). These results are in line with the results of the saturate meta-regression model reported in Table 2.

Compared to M1, the other two best-fitting models, i.e. M2 and M3, are similar except for the inclusion and/or exclusion of a few variables, namely the market platform (included in M2), seller reputation operationalized in terms of the percentage of positive ratings (included in M3), and IV being reputation scores (excluded in M3). The regression coefficients of the variables included in all three models do not show any substantial differences across the three models.

To better evaluate the importance of each moderator, we examine their relative importance values with the method of multi-model inference. As is shown in Table 4, the most important moderators are product usage condition (0.99), the log-transformed number of observations (0.98), DV being final price (0.97) and model being multi-level model (0.95). The relative importance scores of these variables suggest that they are included in almost all models with high weights. Less important moderators are the IV being the reputation score (0.56), whether seller reputation is a percentage (0.41), market platform (0.39), the log-transformed number of parameters (0.39) and whether seller reputation is log-transformed (0.32). This relative importance ranking is consistent with our model selection analysis in that the best-fitting model, M1, includes the five moderators with the highest importance values.

Results on Subset 2

When combining selling probability and selling volume as the dependent variable in Subset 2, the best-fitting is model M4 in Table 5. Model M4 includes region, seller reputation being a percentage rather than an absolute value and the IV being the reputation score rather than the number of positive ratings. As for the region, compared to the USA, seller reputation effects are significantly higher in China (coef. = 0.11, $p < 0.001$). That is, for online transactions in China, the seller reputation scores and positive ratings have larger positive effects on selling probability and selling volume than in the USA. In Europe, the reputation effect is smaller than in the USA but statistically insignificant (coef. = -0.07, $p = 0.15$). Moreover, as expected, the

Table 5 Best-fitting model and relative importance of moderators for Subset 2 focusing on the effect of measures of positive reputation on sales ($n = 117$)

| | M4 | | Relative importance |
|-----------------------------------|----------|------|---------------------|
| | Coef | SE | |
| Const | 0.13*** | 0.04 | |
| <i>Contextual moderators</i> | | | |
| Region | | | 0.96 |
| USA | (ref.) | | |
| China | 0.11*** | 0.03 | |
| Europe | -0.07 | 0.05 | |
| Platform | excl | | 0.11 |
| Year of data | excl | | 0.25 |
| <i>Product-related moderators</i> | | | |
| Condition | excl | | 0.09 |
| Avg. price | excl | | <0.01 |
| <i>Method-related moderators</i> | | | |
| Log-transformed SR | excl | | 0.31 |
| Percentage SR | -0.14*** | 0.05 | 0.92 |
| Product homo | excl | | 0.26 |
| Multilevel | excl | | 0.25 |
| Log (N) | excl | | 0.25 |
| Log (parameter) | excl | | 0.29 |
| <i>Controls</i> | | | |
| Published | excl | | 0.27 |
| Transaction type | excl | | 0.12 |
| DV selling probability | excl | | 0.25 |
| IV reputation score | -0.11** | 0.03 | 0.94 |
| n | 117 | | |
| Log likelihood | 55.59 | | |
| AICc | -98.4 | | |
| Weight | 0.059 | | |

Model M4 is the best-fitting model. The second-best-fitting model is more than 1.5 AICc points away from M4

reputation effect is smaller if seller reputation is measured in percentage than in absolute terms (coef. = -0.14 , $p < 0.001$). That is, among the studies using positive ratings as the independent variable, the effect of percentage positive ratings (e.g., seller with 98% of positive ratings among all ratings) is smaller than the effect of the absolute number of positive ratings (e.g., seller receiving 50 positive ratings).

With regard to the relative importance (see last column in Table 5), the most important moderators are region (0.96), the IV being the reputation score rather than the number of positive ratings (0.94), and seller reputation being a percentage of positive ratings rather than the absolute number of positive ratings (0.92). And these are also the moderators included in the best-fitting model M4.

Discussion

This study builds on and extends a previous meta-analysis that synthesizes evidence from over one hundred empirical studies of the reputation effect in peer-to-peer online markets [30]. The aim of this paper is to explain excess variation in seller reputation effects. Combining model selection analyses and multi-model inference in a meta-analytic context, we find that the excess variation of the observed reputation effects can partly be explained by contextual moderators, product-related moderators and method-related moderators.

In terms of contextual moderators, our results show that the region factor is an important moderator of the effect of measures of positive reputation (i.e. the number of positive ratings and the reputation score) on sales (i.e. Subset 2). Measures of positive reputation have a larger positive effect on the probability of sale and selling volume in studies that use data collected in China rather than the USA and Europe. However, there is no sufficient evidence for a moderating effect of region on the effect of measures of positive reputation on selling price (i.e. Subset 1). One plausible explanation for this observation is that reputable sellers in China or Chinese online platforms (e.g., Taobao) achieve better selling performance through more sales instead of higher prices, suggesting different reputation mechanisms for different types of selling performance. Other contextual moderators are not significant.

For product-related moderators, we only find that used products exhibit a significantly negative moderating effect on the effect of measures of positive reputation on selling price as compared to new products. This is contrary to the expectation that a good reputation is more important for transactions with used products because the uncertainty regarding the condition of these products is higher. One explanation for this finding might be that used products are often relatively unique. For example, for collectors' items there are often very few sellers. This might increase competition between buyers as well as make transactions between specific buyers and sellers more recurrent. Both these mechanisms would lead to reduce the importance of seller reputation. Another, partly related explanation could be that buyers of used products are more risk taking and less attentive to information about seller reputation than buyers of new products. More generally, the moderating effect of product condition might indicate the presence of interaction effects between product, seller and buyer characteristics that could be investigated in future research.

Concerning method-related moderators, we do not find consistent moderators across the two subsets. In Subset 1, multilevel models (i.e. models accounting for clustered observations) produce significantly larger reputation effects whereas studies with larger numbers of observations produce significantly smaller reputation effects. However, we have to be careful to interpret the relation between effect sizes and methodological choices as indications for better (or worse) methodological approaches. Rather, the use of multilevel methodology might be an indication of more sophisticated data collections and better controls for correlated groups. This conjecture could be tested through a re-analyses of studies that used multilevel techniques [5, 12, 56]. The larger N effect is surprising and can certainly not be seen as a reason for smaller N studies, but is also not completely unknown in the literature [59]. Publication bias favoring studies with significant effects could be a reason for

this finding given that smaller N studies are less likely to detect significant effects if effects are smaller. However, also other reasons such as study quality might be behind this finding given that study quality is also related to the size of a study [59].

The smaller effect found on selling volume if the seller's reputation is measured as a percentage of positive ratings rather than in absolute terms was expected. This is an indication that the number of positive ratings is a better indicator for the actual reputation of the seller than the percentage of positive ratings.

Applying model selection and multi-model inference analyses in the meta-analytic context in the current study allows us to present the best-fitting models among the thousands of possible moderator combinations, and evaluate which are the more important moderators in general. However, there are also limitations to this approach. We can only investigate moderators for which enough information is provided in the studies that have been conducted. Therefore, we should also not too strongly interpret the absence of moderation effects of some moderators, because they can also be due to limited power in our dataset. Still, some moderators have a bearing on the size of the reputation effect. However, this study cannot dive deep into the specific mechanisms that underlie these moderating effects, because of the limited information available for each individual study. Dedicated studies that consciously look at variations in certain moderators are necessary to investigate in more detail the different causal explanation behind the moderations [5, 12, 56].

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s42001-022-00160-0>.

Funding R.J. acknowledges financial support by China Scholarship Council (Grant No. 201707720047).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Ahrne, G., Aspers, P., & Brunsson, N. (2015). The organization of markets. *Organization Studies*, 36(1), 7–27. <https://doi.org/10.1177/0170840614544557>
2. Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Akademiai Kiado.
3. Akerlof, G. A. (1970). The market for “lemons”: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3), 488–500. <https://doi.org/10.1016/B978-0-12-214850-7.50022-X>
4. Andrews, T., & Benzing, C. (2007). The determinants of price in internet auctions of used cars. *Atlantic Economic Journal*, 35(1), 43–57. <https://doi.org/10.1007/s11293-006-9045-7>

5. Auspurg, K., & Brüderl, J. (2021). Has the credibility of the social sciences been credibly destroyed? Reanalyzing the “many analysts, one data set” project. *Socius*. <https://doi.org/10.1177/23780231211024421>
6. Auspurg, K., Schneck, A., & Hinz, T. (2019). Closed doors everywhere? A meta-analysis of field experiments on ethnic discrimination in rental housing markets. *Journal of Ethnic and Migration Studies*, 45(1), 95–114. <https://doi.org/10.1080/1369183X.2018.1489223>
7. Barton, K. (2020). *Mu-Min: Multi-model inference*. R Package Version 1.43.17. <http://R-Forge.R-project.org/projects/mumin/>.
8. Beckert, J. (2009). The social order of markets. *Theory and Society*, 38(3), 245–269. <https://doi.org/10.1007/s11186-008-9082-0>
9. Berggren, N., & Jordahl, H. (2006). Free to trust: Economic freedom and social capital. *Kyklos*, 59(2), 141–169. <https://doi.org/10.1111/j.1467-6435.2006.00324.x>
10. Bolton, G. E., Greiner, B., & Ockenfels, A. (2013). Engineering trust: Reciprocity in the production of reputation information. *Management Science*, 59(2), 265–285. <https://doi.org/10.1287/mnsc.1120.1609>
11. Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. Wiley.
12. Breznau, N., Rinke, E., Wuttke, A., Adem, M., Adriaans, J., Alvarez-Benjumea, A., ... Nguyen, H. H. V. (2021). *Observing Many Researchers Using the Same Data and Hypothesis Reveals a Hidden Universe of Uncertainty*. <https://doi.org/10.31222/osf.io/cd5j9>.
13. Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.
14. Buskens, V. & Raub, W. (2013). Rational choice research on social dilemmas: Embeddedness effects on trust. In Wittek, R., Snijders, T.A., Nee, V. (eds.) *The Handbook of Rational Choice Social Research* (pp. 113–150). Stanford University Press. <https://doi.org/10.1515/9780804785501-006>.
15. Cheng, B. S., Altieri, A. H., Torchin, M. E., & Ruiz, G. M. (2019). Can marine reserves restore lost ecosystem functioning? A global synthesis. *Ecology*. <https://doi.org/10.1002/ecy.2617>
16. Cooper, J. D., Han, S. Y. S., Tomasik, J., Ozcan, S., Rustogi, N., van Beveren, N. J., Leweke, F. M., & Bahn, S. (2019). Multimodel inference for biomarker development: An application to schizophrenia. *Translational Psychiatry*, 9(1), 1–10. <https://doi.org/10.1038/s41398-019-0419-4>
17. Cui, X., & Huang, J. (2010). Xinyong Pingjia Tixi yiji Xiangguanyinsu dui Yikoujia Wangshangji-aoyi Yingxiang de Shizhengyanjiu [Empirical study on the impact of the reputation system on buy-it-now online transaction]. *Guanlixuebao*, 7(1), 50–63.
18. Dellarocas, C., & Wood, C. A. (2008). The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science*, 54(3), 460–476. <https://doi.org/10.1287/mnsc.1070.0747>
19. Diekmann, A., Jann, B., Przepiorka, W., & Wehrli, S. (2014). Reputation formation and the evolution of cooperation in anonymous online markets. *American Sociological Review*, 79(1), 65–85. <https://doi.org/10.1177/0003122413512316>
20. Diekmann, A., & Przepiorka, W. (2019). Trust and reputation in markets. In F. Giardini & R. Wittek (Eds.), *The Oxford handbook of gossip and reputation* (pp. 383–400). Oxford University Press.
21. Dini, F., & Spagnolo, G. (2009). Buying reputation on eBay: Do recent changes help? *International Journal of Electronic Business*, 7(6), 581–598. <https://doi.org/10.1504/IJEB.2009.029048>
22. Doleac, J. L., & Stein, L. C. D. (2013). The visible hand: Race and online market outcomes. *Economic Journal*, 123(572), F469–F492. <https://doi.org/10.1111/ecej.12082>
23. Eddhir, A. (2009). *The Value of Reputation in Online Auctions: Evidence from eBay* (Publication No. 1465668) [Master Thesis, Clemson University]. ProQuest Dissertations Publishing.
24. Elfenbein, D. W., Fisman, R., & McManus, B. (2012). Charity as a substitute for reputation: Evidence from an online marketplace. *Review of Economic Studies*, 79(4), 1441–1468. <https://doi.org/10.1093/restud/rds012>
25. Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4), 1645–1692. <https://doi.org/10.1093/qje/qjy013>
26. Frey, V., & Van de Rijt, A. (2016). Arbitrary inequality in reputation systems. *Scientific Reports*, 6, 38304. <https://doi.org/10.1038/srep38304>
27. Hak, T., van Rhee, H & Suurmond, R. (2018). *How to interpret results of meta-analysis*. <https://doi.org/10.2139/ssrn.3241367>. Accessed May 2021.

28. Holper, L. (2020). Raising placebo efficacy in antidepressant trials across decades explained by small-study effects: A meta-reanalysis. *Frontiers in Psychiatry, 11*, 633. <https://doi.org/10.3389/fpsy.2020.00633>
29. Jian, L., Yang, S., Ba, S., Lu, L., & Jiang, L. C. (2019). Managing the crowds: The effect of prize guarantees and in-process feedback on participation in crowdsourcing contests. *MIS Quarterly, 43*(1), 97–112. <https://doi.org/10.25300/MISQ/2019/13649>
30. Jiao, R., Przepiorka, W., & Buskens, V. (2021). Reputation effects in peer-to-peer online markets: A meta-analysis. *Social Science Research, 95*, 102522. <https://doi.org/10.1016/j.ssresearch.2020.102522>
31. Kollock, P. (1999). The production of trust in online markets. In E. J. Lawler, M. Macy, S. Thyne, & H. A. Walker (Eds.), *Advances in group processes* (Vol. 16, pp. 99–123). JAI Press.
32. Lavoie, R. A., Amyot, M., & Lapierre, J. F. (2019). Global meta-analysis on the relationship between mercury and dissolved organic carbon in freshwater environments. *Journal of Geophysical Research: Biogeosciences, 124*(6), 1508–1523. <https://doi.org/10.1029/2018JG004896>
33. Lindenberg, S., Wittek, R., & Giardini, F. (2020). Reputation effects, embeddedness, and Granovetter's error. In V. Buskens, R. Corten, & C. Snijders (Eds.), *Advances in the sociology of trust and cooperation* (pp. 113–140). De Gruyter.
34. Liu, Y. W., Chen, H. P., Wie, G. J., & Xu, J. L. (2007). Huicui fenxi: Xinyong pingjia neng cujin wangshangpaimai ma [Does reputation system impact online auction results: A meta-analysis]. *Xinxi Xitong Xuebao, 1*(1), 16–33.
35. Ljunge, M. (2014). Social capital and political institutions: Evidence that democracy fosters trust. *Economics Letters, 122*(1), 44–49. <https://doi.org/10.1016/j.econlet.2013.10.031>
36. Lo Iacono, S., & Quaranta, M. (2019). Contextual economic conditions, institutions and social trust: Trends and cross-national differences in Europe, 2002–2017. *Polis, 33*(2), 185–214. <https://doi.org/10.1424/94245>
37. Lo Iacono, S., & Sonmez, B. (2021). The effect of trusting and trustworthy environments on the provision of public goods. *European Sociological Review, 37*(1), 155–168. <https://doi.org/10.1093/esr/jcaa040>
38. Matsuki, K., Kuperman, V., & van Dyke, J. A. (2016). The random forests statistical technique: An examination of its value for the study of reading. *Scientific Studies of Reading, 20*(1), 20–33. <https://doi.org/10.1080/10888438.2015.1107073>
39. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
40. Nannestad, P. (2008). What have we learned about generalized trust, if anything? *Annual Review of Political Science, 11*, 413–436. <https://doi.org/10.1146/annurev.polisci.11.060606.135412>
41. Nee, V. (2005). The new institutionalisms in economics and sociology. In N. J. Smelser & R. Swedberg (Eds.), *The handbook of economic sociology* (pp. 49–74). Princeton University Press.
42. Nosko, C., & Tadelis, S. (2015). *The limits of reputation in platform markets: An empirical analysis and field experiment* (No. w20830). National Bureau of Economic Research. <https://doi.org/10.3386/w20830>
43. Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *JAMA, 295*(6), 676–680. <https://doi.org/10.1001/jama.295.6.676>
44. Podolny, J. M. (1994). Market uncertainty and the social character of economic exchange. *Administrative Science Quarterly, 39*(1), 1–63. <https://doi.org/10.2307/2393299>
45. Przepiorka, W. (2011). Ethnic discrimination and signals of trustworthiness in anonymous online markets: Evidence from two field experiments. *Zeitschrift für Soziologie, 40*(2), 132–141.
46. Przepiorka, W. (2013). Buyers pay for and sellers invest in a good reputation: More evidence from eBay. *Journal of Socio-Economics, 42*, 31–42. <https://doi.org/10.1016/j.socec.2012.11.004>
47. Przepiorka, W., & Berger, J. (2017). Signalling theory evolving: Signals and signs of trustworthiness in social exchange. In B. Jann & W. Przepiorka (Eds.), *Social dilemmas, institutions and the evolution of cooperation* (pp. 373–392). De Gruyter Oldenbourg.
48. Przepiorka, W., Norbutas, L., & Corten, R. (2017). Order without law: Reputation promotes cooperation in a cryptomarket for illegal drugs. *European Sociological Review, 33*(6), 752–764. <https://doi.org/10.1093/esr/jcx072>
49. Putnam, R. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton University Press.
50. Resnick, P., Kuwabara, K., Zeckhauser, R., & Friedman, E. (2000). Reputation systems. *Communications of the ACM, 43*(12), 45–48.

51. Roth, A. E. (2015). *Who gets what—and why: Understand the choices you have improve the choices you make*. William Collins.
52. Samia, D. S., Bessa, E., Blumstein, D. T., Nunes, J. A., Azzurro, E., Morroni, L., Sbragaglia, V., Januchowski-Hartley, F. A., & Geffroy, B. (2019). A meta-analysis of fish behavioural reaction to underwater human presence. *Fish and Fisheries*, 20(5), 817–829. <https://doi.org/10.1111/faf.12378>
53. Schilke, O., Reimann, M., & Cook, K. S. (2021). Trust in social relations. *Annual Review of Sociology*. <https://doi.org/10.1146/annurev-soc-082120-082850>
54. Schlägel, C. (2011). *Country-specific effects of reputation: A cross-country comparison of online auction markets*. Springer.
55. Shapiro, C. (1983). Premiums for high quality products as returns to reputations. *The Quarterly Journal of Economics*, 98(4), 659–679. <https://doi.org/10.2307/1881782>
56. Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š, Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., & Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*. <https://doi.org/10.1177/2515245917747646>
57. Snijders, C., & Zijdeman, R. (2004). Reputation and internet auctions: eBay and beyond. *Analyse & Kritik*, 26(1), 158–184. <https://doi.org/10.1515/auk-2004-0109>
58. Steinhardt, H. C. (2012). How is high trust in China possible? Comparing the origins of generalized trust in three Chinese societies. *Political Studies*, 60(2), 434–454. <https://doi.org/10.1111/j.1467-9248.2011.00909.x>
59. Sterne, J. A., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology*, 53(11), 1119–1129. [https://doi.org/10.1016/S0895-4356\(00\)00242-0](https://doi.org/10.1016/S0895-4356(00)00242-0)
60. Swamynathan, G., Almeroth, K. C., & Zhao, B. Y. (2010). The design of a reliable reputation system. *Electronic Commerce Research*, 10(3), 239–270. <https://doi.org/10.1007/s10660-010-9064-y>
61. Symonds, M. R., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, 65(1), 13–21. <https://doi.org/10.1007/s00265-010-1037-6>
62. Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics*, 8, 321–340. <https://doi.org/10.1146/annurev-economics-080315-015325>
63. Tan, S. J., & Tambyah, S. K. (2011). Generalized trust and trust in institutions in Confucian Asia. *Social Indicators Research*, 103(3), 357–377. <https://doi.org/10.1007/s11205-010-9703-7>
64. Uslaner, E. M. (2002). *The moral foundations of trust*. Cambridge University Press.
65. Uzzi, B. (1996). The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American Sociological Review*. <https://doi.org/10.2307/2096399>
66. Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
67. Yin, H. (2017). Wangluojiaoyi Zhong Xinyujiazhi de Yingxiangyinsu Yanjiu—Jiyu Taobaowang de Shizhengfenxi [Affect factors of reputation value on internet transaction—An empirical analysis based on taobao]. *Shangyeyingji yu Guanli*, 309, 16–28.
68. Young, C., & Holsteen, K. (2017). Model uncertainty and robustness: A computational framework for multimodel analysis. *Sociological Methods & Research*, 46(1), 3–40. <https://doi.org/10.1177/0049124115610347>
69. Zhao, J., & Huang, J. (2008). An empirical research on Taobao: Seller reputation's impact on auction price premium. *IEEE Symposium on Advanced Management of Information for Globalized Enterprises*. <https://doi.org/10.1109/AMIGE.2008.ECP.52>
70. Zhou, G. (2014). Buwanquanxinxi xiade Jiage Xinhao Boyi: Laizi Taobaowang de Zhengju [Price signaling game in the environment of incomplete information: Evidence from taobao.com]. *Dang-daicaijing*, 4, 14–23.
71. Zhu, Y., Li, Y., & Leboulanger, M. (2009). National and cultural differences in the C2C electronic marketplace: An investigation into transactional behaviors of Chinese, American, and French consumers on eBay. *Tsinghua Science and Technology*, 14(3), 383–389. [https://doi.org/10.1016/S1007-0214\(09\)70055-0](https://doi.org/10.1016/S1007-0214(09)70055-0)